ω-rational Languages: High Complexity Classes vs. Borel Hierarchy

Enrico Formenti, Markus Holzer, Martin Kutrib, Julien Provillard

To cite this version:

Enrico Formenti, Markus Holzer, Martin Kutrib, Julien Provillard. ω-rational Languages: High Complexity Classes vs. Borel Hierarchy. Language and Automata Theory and Applications, Mar 2014, Madrid, Spain. 10.1007/978-3-319-04921-2_30 . hal-01297613

HAL Id: hal-01297613

https://hal.science/hal-01297613

Submitted on 4 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ω-rational languages: high complexity classes vs. Borel Hierarchy*

Enrico Formenti[1], Markus Holzer[2], Martin Kutrib[2], and Julien Provillard[1]

[1] Université Nice Sophia Antipolis,
CNRS, I3S, UMR 7271, 06900 Sophia Antipolis (France).
{enrico.formenti,julien.provillard}@unice.fr

[2] Justus-Liebig Universität Gießen, Institut für Informatik,
Arndtstraße 2, 35392 Gießen (Germany).
{markus.holzer, martin.kutrib}@informatik.uni-giessen.de

Abstract. The paper investigates classes of languages of infinite words with respect to the acceptance conditions of the finite automata recognizing them. Some new natural classes are compared with the Borel hierachy. In particular, it is proved that $(\texttt{fin}, =)$ is as high as $\mathsf{F}_\sigma^\mathsf{R}$ and $\mathsf{G}_\delta^\mathsf{R}$. As a side effect, it is also proved that in this last case, considering or not considering the initial state of the FA makes a substantial difference.

Keywords: ω-rational languages, Borel hierarchy, acceptance conditions

1 Introduction

Languages over infinite words have been used since the very introduction of symbolic dynamics. Afterwards, they have spread in a multitude of scientific fields. Computer science is more directly concerned for example by their application in formal specification and verification, game theory, logics, *etc.*.

ω-rational languages have been introduced as a natural extension of languages of finite words recognized by finite automata. Indeed, a finite automaton accepts some input u if at the end of the reading of u, the automaton reaches a final state. Clearly, when generalizing to infinite words, this accepting condition has to be changed. For this reason, new accepting conditions have been introduced in literature. For example, an infinite word w is accepted by a finite automaton $\mathcal{A}$ under the Büchi acceptance condition if and only if there exists a run of $\mathcal{A}$ which passes infinitely often through a set of accepting states while reading w. Indeed, this was introduced by Richard Büchi in the seminal work [1] in 1960.

Later on, David Muller characterized runs that pass through all elements of a given set of accepting states and visit them infinitely often [8]. Afterwards, more acceptance conditions appeared in a series of papers [4,5,11,7,6]. Each of these

* This work has been partially supported by the French National Research Agency project EMC (ANR-09-BLAN-0164)

works was trying to capture a particular semantic on the runs or to fill some conceptual gap. Acceptance conditions are selectors for runs of the automaton under consideration. Of course, the set of selected runs is also deeply influenced by the structural properties of the FA: deterministic vs. non-deterministic, complete vs. non complete (see for instance [6]).

Each acceptance condition characterizes a class of languages. In [2], it is proved that if the acceptance condition is definable in MSO (monadic second order) logic then the class of languages it induces is ω-rational. However, more work was necessary to find which was the overall picture *i.e.* which are the relations between classes of languages induced by the acceptance conditions appeared in literature so far. The well-known Borel hierarchy constitute the backbone of such a picture. Classes in the hierarchy are ordered by set inclusion.

This paper continues the classification work closing some open questions concerning the positioning of the class of languages induced by CDFA$(\mathtt{fin}, =)$ (*i.e.* languages characterized by runs that pass finitely many times through all the elements of a given set of final states, recognized by Complete Deterministic Finite Automata). The motivation for a further study of the condition $(\mathtt{fin}, =)$ is twofold. From one hand, this class is, in a sense, surprising. Indeed, it is as high as the highest classes of the Borel hierachy but it is distinct from them. The interest of such a result is to have examples of languages that have high complexity but in which the complexity is not just determined by the topology one defines over the words (the Cantor topology here) but the complexity is determined by the intrinsic combinatorial complexity of the words themselves.

From the other hand, it is another step in the understanding of the theory of formal specification and verification of daemon processes (non-terminating processes). In this case, a run of the process is accepted only if it passes through a finite number of exceptions.

The paper also highlights an interesting phenomenon: the complexity class can be greatly influenced by the fact that one considers the very first elements of the paths (initial node) or not. In the sequel given an acceptance condition (c, R), the version in which the initial node is considered is denoted (c', R).

For example $\mathtt{run}$ is the set of states visited by the finite automaton while reading the input word, excluding the initial state; $\mathtt{run}'$ is the same as $\mathtt{run}$ but includes the initial state. By Proposition 21, one finds that CDFA$(\mathtt{fin}, =) \subsetneq$ CDFA$(\mathtt{fin}', =)$ (CDFA stands for complete deterministic finite automata). As a consequence CDFA$(\mathtt{fin}', =)$ is even higher than CDFA$(\mathtt{fin}, =)$. The rest of the paper is devoted in proving (or disproving) the inclusion relations *wrt.* all previously known classes. The resulting hierarchy is illustrated in Figure 5.

Most of the proofs have been omitted due to a lack of space. They will appear in the long version of this article.

2 Languages and automata

Let $\mathbb{N}$ denote the set of non-negative integers. For all $i, j \in \mathbb{N}$, $[i, j]$ is the set $\{i, i+1, \ldots, j\}$. For a set A, $|A|$ denotes the cardinality of A and $\mathcal{P}(A)$ the powerset of A. An *alphabet* is a finite set and a *letter* is an element of an alphabet. Given an alphabet Σ, a *word* over Σ is a sequence of letters from Σ. Let Σ^* and Σ^ω denote the set of all finite words and the set of all infinite words over Σ, respectively. Let Σ^∞ denote $\Sigma^* \cup \Sigma^\omega$. For a word u, $|u|$ denotes the length of u and $|u|_a$ denotes the number of occurrences of the letter a in u. The empty word ε is the only word of length zero. For all words $u \in \Sigma^*$ and $v \in \Sigma^\infty$, uv denotes the concatenation of u with v. For all word $u \in \Sigma^\infty$, for all $0 \leq i \leq j < |u|$, the word $u_i u_{i+1} \dots u_j$ is denoted by $u_{[i,j]}$.

A *language* is a subset of Σ^*, similarly an ω-*language* is a subset of Σ^ω. For a language $\mathcal{L}_1$ and for $\mathcal{L}_2 \in \Sigma^\infty$, $\mathcal{L}_1\mathcal{L}_2 = \{uv \in \Sigma^* : u \in \mathcal{L}_1, v \in \mathcal{L}_2\}$ denotes the concatenation of $\mathcal{L}_1$ with $\mathcal{L}_2$. For a language $\mathcal{L} \subseteq \Sigma^*$, let $\mathcal{L}^0 = \{\varepsilon\}$, $\mathcal{L}^{n+1} = \mathcal{L}^n\mathcal{L}$ and $\mathcal{L}^* = \bigcup_{n\in\mathbb{N}} \mathcal{L}^n$ the Kleene star of $\mathcal{L}$. For a language $\mathcal{L}$, the infinite iteration of $\mathcal{L}$ is the ω-language $\mathcal{L}^\omega = \{u_0u_1u_2 \cdots : \forall i \in \mathbb{N}, u_i \in \mathcal{L} \setminus \{\epsilon\}\}$.

The class of *rational languages* is the smallest class of languages containing $\emptyset$, all sets $\{a\}$ (for $a \in \Sigma$) and which is closed under union, concatenation and Kleene star operations. An ω-language $\mathcal{L}$ is ω-*rational* if there exist $n \in \mathbb{N}$ and two families $\{\mathcal{L}_i\}$ and $\{\mathcal{L}'_i\}$ of n rational languages such that $\mathcal{L} = \bigcup_{i=0}^{n-1} \mathcal{L}'_i\mathcal{L}_i^\omega$. Let RAT denote the set of all ω-rational languages.

Rational languages and ω-rational languages are denoted by rational expressions. For instance, for the alphabet $\Sigma = \{0,1\}$, Σ^*1 denotes the language of words ending with a 1 while $(\Sigma^*1)^\omega$ and $\Sigma^*(0^\omega + 1^\omega)$ denote the ω-languages of words containing an infinite number of 1's, and a finite number of 0's or a finite number of 1's, respectively.

A *finite automaton* (FA) is a tuple $(\Sigma, Q, T, I, \mathcal{F})$ where Σ is an alphabet, Q a finite set of states, $T \subseteq Q \times \Sigma \times Q$ is the set of *transitions*, $I \subseteq Q$ is the set of *initial states* and $\mathcal{F} \subseteq \mathcal{P}(Q)$ is the *acceptance table*. A FA is a *deterministic* finite automaton (DFA) if $|I| = 1$ and $|\{q \in Q : (p,a,q) \in T\}| \leq 1$ for all $p \in Q$, $a \in \Sigma$. It is a *complete* finite automaton (CFA) if $|\{q \in Q : (p,a,q) \in T\}| \geq 1$ for all $p \in Q$, $a \in \Sigma$. A CDFA is a FA which is both deterministic and complete.

A CDFA induces a transition function $\delta : Q \times \Sigma \to Q$ such that for all $p \in Q$ and $a \in \Sigma$, $\delta(p,a)$ is the only state such that $(p, a, \delta(p,a)) \in T$. The transition function can be extended to a function $\delta' : Q \times \Sigma^* \to Q$ by defining for all $p \in Q$, $\delta'(p,\varepsilon) = p$ and for all $p \in Q$, $a \in \Sigma$ and $u \in \Sigma^*$, $\delta'(p, au) = \delta'(\delta(p,a), u)$. We usually make no distinction between δ and δ'.

If $I = \{q_0\}$ for some state $q_0 \in Q$, we shall write $(\Sigma, Q, T, q_0, \mathcal{F})$ instead of $(\Sigma, Q, T, I, \mathcal{F})$. Similarly, if $\mathcal{F} = \{F\}$ or $\mathcal{F} = \{\{f\}\}$, we shall write (Σ, Q, T, I, F) or (Σ, Q, T, I, f) instead of $(\Sigma, Q, T, I, \mathcal{F})$, respectively.

An *infinite path* in a FA $\mathcal{A} = (\Sigma, Q, T, I, \mathcal{F})$ is a sequence $(p_i, x_i)_{i\in\mathbb{N}}$ such that $(p_i, x_i, p_{i+1}) \in T$ for all $i \in \mathbb{N}$. The (infinite) word x is the *label* of the path. A *finite path* from p to q is a sequence $(p_i, u_i)_{i\in[0,n]}$ for some n such that $p_0 = p$, for all $i \in [0, n-1]$, $(p_i, u_i, p_{i+1}) \in T$ and $(p_n, u_n, q) \in T$. The (finite) word u is the label of the path. A path is *initial* if $p_0 \in I$. A state q is *accessible* if there exists an initial path to q and $\mathcal{A}$ is *accessible* if all its states are. A *loop* is a path from a state to the same state. The FA $\mathcal{A}$ is *normalized* if it is accessible, $I = \{q_0\}$ for some $q_0 \in Q$ and q_0 does not belong to a loop.

3 Acceptance conditions, classes of languages and topology

Definition 1. *Let $\mathcal{A} = (\Sigma, Q, T, I, \mathcal{F})$ be a* FA *and $p = (p_i, x_i)_{i\in\mathbb{N}}$ a path in $\mathcal{A}$. Define the sets*

- $\mathtt{run}_{\mathcal{A}}(p) = \{q \in Q : \exists i > 0, p_i = q\}$,
- $\mathtt{run}'_{\mathcal{A}}(p) = \{q \in Q : \exists i \geq 1, p_i = q\}$,
- $\mathtt{inf}_{\mathcal{A}}(p) = \{q \in Q : \forall i > 0, \exists j \geq i, p_j = q\}$,
- $\mathtt{fin}_{\mathcal{A}}(p) = \mathtt{run}_{\mathcal{A}}(p) \smallsetminus \mathtt{inf}_{\mathcal{A}}(p)$,
- $\mathtt{fin}'_{\mathcal{A}}(p) = \mathtt{run}'_{\mathcal{A}}(p) \smallsetminus \mathtt{inf}_{\mathcal{A}}(p)$,
- $\mathtt{ninf}_{\mathcal{A}}(p) = Q \smallsetminus \mathtt{inf}_{\mathcal{A}}(p)$

as the sets of states appearing at least one time *(counting or not the first state of the path),* infinitely many times*,* finitely many times but at least once *(counting or not the first state of the path), and either* finitely many times including never *in p, respectively.*

An *acceptance condition* for $\mathcal{A}$ is a subset of all the initial infinite paths of $\mathcal{A}$. The paths inside such a subset are called *accepting paths*. Let $\mathcal{A}$ be a FA and *cond* be an acceptance condition for $\mathcal{A}$, a word x is *accepted* by $\mathcal{A}$ (under condition *cond*) if and only if it is the label of some accepting path.

Let $\sqcap$ be the binary relation over sets such that for all sets A and B, $A \sqcap B$ if and only if $A \cap B \neq \emptyset$.

In this paper, we consider acceptance conditions induced by pairs $(c, \mathbf{R}) \in \{\mathtt{run}, \mathtt{run}', \mathtt{inf}, \mathtt{fin}, \mathtt{fin}', \mathtt{ninf}\} \times \{\sqcap, \subseteq, =\}$. A pair $cond = (c, \mathbf{R})$ defines an acceptance condition $cond_{\mathcal{A}}$ on an automaton $\mathcal{A} = (\Sigma, Q, T, I, \mathcal{F})$ as follows: an initial infinite path $p = (p_i, x_i)_{i\in\mathbb{N}}$ is accepting if and only if there exists a set $F \in \mathcal{F}$ such that $c_{\mathcal{A}}(p)\ \mathbf{R}\ F$. We denote by $\mathcal{L}^{cond}_{\mathcal{A}}$ the *language accepted by $\mathcal{A}$ under the acceptance condition $cond_{\mathcal{A}}$*, *i.e.*, the set of all words accepted by $\mathcal{A}$ under $cond_{\mathcal{A}}$.

Remark 2. For acceptance conditions which use the relation $\sqcap$, we can assume that the acceptance table is reduced to one set of states, taking, if necessary, the union of all sets in the acceptance table.

Definition 3. *For all pairs $cond \in \{\mathtt{run}, \mathtt{run}', \mathtt{inf}, \mathtt{fin}, \mathtt{fin}', \mathtt{ninf}\} \times \{\sqcap, \subseteq, =\}$ and for all finite alphabets Σ, define the following sets*

- $\mathrm{FA}^{(\Sigma)}(cond) = \left\{\mathcal{L}^{cond}_{\mathcal{A}},\ \mathcal{A} \text{ is a FA on } \Sigma\right\}$,
- $\mathrm{DFA}^{(\Sigma)}(cond) = \left\{\mathcal{L}^{cond}_{\mathcal{A}},\ \mathcal{A} \text{ is a DFA on } \Sigma\right\}$,
- $\mathrm{CFA}^{(\Sigma)}(cond) = \left\{\mathcal{L}^{cond}_{\mathcal{A}},\ \mathcal{A} \text{ is a CFA on } \Sigma\right\}$,
- $\mathrm{CDFA}^{(\Sigma)}(cond) = \left\{\mathcal{L}^{cond}_{\mathcal{A}},\ \mathcal{A} \text{ is a CDFA on } \Sigma\right\}$

as the classes of ω-languages on Σ accepted by FA*,* DFA*,* CFA*, and* CDFA*, respectively, under the acceptance condition derived by cond. When it is not confusing, we omit to precise the alphabet in these notations.*

When Σ is endowed with discrete topology and Σ^ω with the induced product topology, let F, G, F_σ and G_δ be the collections of all closed sets, open sets, countable unions of closed set and countable intersections of open sets, respectively. For any pair A, B of collections of sets, denote by $\mathcal{B}(A)$, $A \mathbin{\Delta} B$, and A^{R} the Boolean closure of A, the set $\{U \cap V : U \in A, V \in B\}$ and the set $A \cap \mathsf{RAT}$, respectively. These, indeed, are the lower classes of the Borel hierarchy. For more on this subject we refer the reader to [12] or [9], for instance.

Some of the acceptance conditions derived by pairs $(c, \mathbf{R})$ have been studied in the literature (see [1,8,4,5,11,7,6,10,3]). It is known that all the classes of languages induced are subclasses of RAT because the acceptance conditions are MSO-definable, see [1,2]. The known inclusions are depicted in Figure 5.

In the sequel, we deal with languages sharing the same structure. For an alphabet Σ, $a \in \Sigma$, $k \geq 0$ and $n > 0$, we denote the language

$$\{x \in \Sigma^\omega : |x|_a = k \pmod{n}\}$$

by $\mathcal{L}_{k,n}^{\Sigma,a}$ and $\tilde{\mathcal{L}}_{k,n}^{\Sigma,a}$ denotes the language $\mathcal{L}_{k,n}^{\Sigma,a} + (\Sigma^* a)^\omega$.

4 Some relations between run and run′, and fin and fin′

The following lemma is immediate.

Lemma 4. *Let cond* $\in \{\mathtt{run}, \mathtt{inf}, \mathtt{fin}, \mathtt{ninf}\} \times \{\sqcap, \subseteq, =\}$. *If a language* $\mathcal{L}$ *is recognized by an automaton under condition cond, then it is recognized by a normalized automaton which is complete (resp. deterministic) if the initial one is complete (resp. deterministic) under condition cond.*

Corollary 5. *Let* $(c, \boldsymbol{R}) \in \{\mathtt{run}, \mathtt{fin}\} \times \{\sqcap, \subseteq, =\}$. *The class of languages induced by* $(c, \boldsymbol{R})$ *is included in the respective class of languages induced by* $(c', \boldsymbol{R})$.

Lemma 6. *Let* $\boldsymbol{R} \in \{\sqcap, \subseteq, =\}$ *and cond* $= (\mathtt{run}', \boldsymbol{R})$. *If a language* $\mathcal{L}$ *is recognized by an automaton under condition cond, then it is recognized by a normalized automaton which is complete (resp. deterministic) if the initial one is complete (resp. deterministic) under condition cond.*

Proposition 7. *Let* $\boldsymbol{R} \in \{\sqcap, \subseteq, =\}$. *The conditions* $(\mathtt{run}, \boldsymbol{R})$ *and* $(\mathtt{run}', \boldsymbol{R})$ *induce the same classes of languages.*

We will see later that Proposition 7 has no equivalence for condition based on `fin`. In general, the inclusion of classes induced by `fin` in the respective class induced by `fin′` is strict.

From now on, without loss of generality, we assume that Σ is an alphabet containing $\{0, 1\}$ and we denote the set $\Sigma \setminus \{1\}$ by Σ_0 and the set $\Sigma \setminus \{0\}$ by Σ_1.

5 The acceptance conditions $(\mathtt{fin}, \sqcap)$ and $(\mathtt{fin}', \sqcap)$

The acceptance condition $(\mathtt{fin}, \sqcap)$ has already been studied in [6]. In this paper, we prove that the condition $(\mathtt{fin}', \sqcap)$ defines new classes for deterministic or complete automata.

Proposition 8. *The class* $\mathrm{FA}(\mathtt{fin}', \sqcap)$ *is included in the class* $\mathrm{FA}(\mathtt{fin}, \sqcap)$*.*

Proposition 9. *The language* $\mathcal{L}_{0,2}^{\Sigma,1}$ *is in* $\mathrm{CDFA}(\mathtt{fin}', \sqcap)$ *but not in* $\mathrm{CFA}(\mathtt{fin}, \sqcap)$ *or in* $\mathrm{DFA}(\mathtt{fin}, \sqcap)$*.*

Proof. Remark that $\mathcal{L}_{0,2}^{\Sigma,1} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}', \sqcap)}$ for the CDFA $\mathcal{A} = (\Sigma, \{q_0, q_1\}, T, q_0, q_1)$ where $(p, a, q) \in T$ if and only if $a = 1$ and $p \neq q$ or $a \neq 1$ and $p = q$.

For the sake of argument, assume that $\mathcal{L}_{0,2}^{\Sigma,1} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}, \sqcap)}$ for a CFA $\mathcal{A}$. The word $x = 0^\omega$ is in $\mathcal{L}_{0,2}^{\Sigma,1}$ so there exists an accepting path $p = (p_i, x_i)_{i\in\mathbb{N}}$ in $\mathcal{A}$ under $(\mathtt{fin}, \sqcap)$. Let $k > 0$ such that $p_k \in F$ is visited finitely often in p. Let $y = 0^k10^\omega$, y is not in $\mathcal{L}$, then all paths starting from p_k and labeled by 10^ω visit p_k infinitely often. Therefore, there exists a loop on p_k labeled by $10^{k'}$ for some $k' \in \mathbb{N}$. Inserting this loop one time in the first path, we find an accepting path labeled by y, this is a contradiction.

For the sake of argument, assume that $\mathcal{L}_{0,2}^{\Sigma,1} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}, \sqcap)}$ for a DFA $\mathcal{A}$. Without loss of generality, we can assume that $\mathcal{A}$ is accessible. As for all $u \in \Sigma^*$, $u0^\omega$ or $u10^\omega$ is in $\mathcal{L}_{0,2}^{\Sigma,1}$, there exists a finite initial path labeled by u and $\mathcal{A}$ is complete. We have just shown that this is not possible. □

Theorem 10. *The following relations hold for the classes induced by* $(\mathtt{fin}', \sqcap)$*:*

1. $\mathrm{CDFA}(\mathtt{fin}, \sqcap) \subsetneq \mathrm{CDFA}(\mathtt{fin}', \sqcap)$, $\mathrm{DFA}(\mathtt{fin}, \sqcap) \subsetneq \mathrm{DFA}(\mathtt{fin}', \sqcap)$, $\mathrm{CFA}(\mathtt{fin}, \sqcap) \subsetneq \mathrm{CFA}(\mathtt{fin}', \sqcap)$,
2. $\mathrm{FA}(\mathtt{fin}, \sqcap) = \mathrm{FA}(\mathtt{fin}', \sqcap)$,
3. $\mathrm{CDFA}(\mathtt{fin}', \sqcap) \subsetneq \mathrm{CFA}(\mathtt{fin}', \sqcap) \subsetneq \mathrm{FA}(\mathtt{fin}', \sqcap)$,
4. $\mathrm{CDFA}(\mathtt{fin}', \sqcap) \subsetneq \mathrm{DFA}(\mathtt{fin}', \sqcap) \subsetneq \mathrm{FA}(\mathtt{fin}', \sqcap)$.

There are no other relations for the classes induced by $(\mathtt{fin}', \sqcap)$ *except those obtained by transitivity with previously known classes.*

Proof. The first point follows from Corollary 5 and Proposition 9. The equality $\mathrm{FA}(\mathtt{fin}, \sqcap) = \mathrm{FA}(\mathtt{fin}', \sqcap)$ holds from Corollary 5 and Proposition 8. The incomparability of $\mathrm{DFA}(\mathtt{fin}', \sqcap)$ with $\mathrm{CFA}(\mathtt{fin}', \sqcap)$ and the fact there is no other inclusions come from results of [6]. Indeed, at the one hand, $\mathsf{F}^{\mathsf{R}} \subseteq \mathrm{DFA}(\mathtt{fin}, \sqcap)$ but $\mathsf{F}^{\mathsf{R}} \not\subseteq \mathrm{CFA}(\mathtt{fin}', \sqcap)$. And, at the other hand, the language $\Sigma^*10\Sigma^\omega + \Sigma^*0^\omega$ is in $(\mathrm{CDFA}(\mathtt{ninf}, \sqcap) \cap \mathrm{CFA}(\mathtt{fin}, \sqcap)) \setminus \mathrm{DFA}(\mathtt{fin}', \sqcap)$. Finally, the language Σ^*0^ω is in $\mathrm{CDFA}(\mathtt{fin}, \sqcap) \cap (\mathsf{F}_\sigma^{\mathsf{R}} \setminus \mathsf{G}_\delta^{\mathsf{R}})$. □

6 The acceptance conditions $(\mathtt{fin}, \subseteq)$ and $(\mathtt{fin}', \subseteq)$

In [3], it is proved that an automaton using the acceptance condition $(\mathtt{fin}, \subseteq)$ and $(\mathtt{fin}, =)$ can be completed without changing the recognized language. It follows that the completeness does not matter for classes induced by those conditions. The same holds for $(\mathtt{fin}', \subseteq)$ and $(\mathtt{fin}', =)$.

Proposition 11. *The class* F *is included in* $\mathrm{CDFA}(\mathtt{fin}, \subseteq)$ *and the class* $\mathsf{F}_\sigma \cap \mathsf{G}_\delta$ *is included in* $\mathrm{CDFA}(\mathtt{fin}, =)$.

Proposition 12 ([2]). *The class* $\mathrm{CDFA}(\mathtt{fin}', \subseteq)$ *is included in* G_δ.

Proposition 13. *The language* $(\Sigma^* 1)^\omega$ *is in* $\mathrm{CDFA}(\mathtt{fin}, \subseteq) \setminus \mathsf{F}_\sigma^{\mathsf{R}}$.

Lemma 14. *Let* $\mathcal{L}$ *be a language in* $\mathrm{FA}(\mathtt{fin}, \subseteq)$ *(resp. in* $\mathrm{FA}(\mathtt{fin}', \subseteq)$*) such that there exists* $a, b \in \Sigma$, $u \in \Sigma^*$ *and for all* $k \in \mathbb{N}$, $ba^k u a^\omega \in \mathcal{L}$ *(resp.* $a^k u a^\omega \in \mathcal{L}$*). Then* ba^ω *(resp.* a^ω*) is in* $\mathcal{L}$.

Proof. Let $\mathcal{A} = (\Sigma, Q, T, I, \mathcal{F})$ such that $\mathcal{L} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}, \subseteq)}$ (resp. $\mathcal{L} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}', \subseteq)}$). Let $n = |Q|$, as $x = ba^n u a^\omega$ (resp. $x = a^n u a^\omega$) is in $\mathcal{L}$, there exists an accepting path $p = (p_i, x_i)_{i \in \mathbb{N}}$ in $\mathcal{A}$. There exists k, k' such that $1 \le k < k' \le n+1$ (resp. $0 \le k < k' \le n$) and $p_k = p_{k'}$. Choose k minimal. We define a path $p' = (p'_i, y_i)_{i \in \mathbb{N}}$ in $\mathcal{A}$ where $y = ba^\omega$ (resp. $y = a^\omega$), for all $i \in [0, k], p'_i = p_i$ and for all $i \in \mathbb{N}, p'_{k+i} = p_{k + (i \pmod{k'-k})}$. If p' is accepting, we can conclude. If not, then, by minimality of k, $\mathtt{fin}(p') = \{p_i : i \in [1, k-1]\}$ (resp. $\mathtt{fin}'(p') = \{p_i : i \in [0, k-1]\}$) is not included in any $F \in \mathcal{F}$. But as p is accepting, there exists $F \in \mathcal{F}$ such that $\mathtt{fin}(p) \subseteq F$ (resp. $\mathtt{fin}'(p) \subseteq F$). That means there exists $q \in \mathtt{fin}(p')$ (resp. $q \in \mathtt{fin}'(p')$) such that $q \in \mathtt{inf}(p)$. Let $k_0 \in [1, k-1]$ (resp. $k_0 \in [0, k-1]$) be minimal such that $p_{k_0} \in \mathtt{inf}(p)$. Then by definition of $\mathtt{inf}(p)$, we can find an index k'_0 such that $p_{k'_0} = p_{k_0}$, $k'_0 \ge |u| + n + 1$ and for all $i \ge k'_0, p_i \in \mathtt{inf}(p)$. We define a path $p'' = (p''_i, y_i)_{i \in \mathbb{N}}$ in $\mathcal{A}$ where for all $i \in [0, k_0], p''_i = p_i$ and for all $i \in \mathbb{N}, p'_{k_0+i} = p_{k'_0+i}$. By minimality of k_0 and by definition of k'_0, $\mathtt{fin}(p'') = \{p_i : i \in [1, k_0 - 1]\} \subseteq \mathtt{fin}(p) \subseteq F$ (resp. $\mathtt{fin}'(p'') = \{p_i : i \in [0, k_0 - 1]\} \subseteq \mathtt{fin}'(p) \subseteq F$) and p'' is an accepting path labeled by y. ☐

Proposition 15. *The language* $\Sigma^* 1 \Sigma^\omega$ *is in* $\mathrm{CDFA}(\mathtt{ninf}, \sqcap) \cap \mathsf{G}^{\mathsf{R}} \setminus \mathrm{FA}(\mathtt{fin}', \subseteq)$.

Proposition 16. *The language* $\tilde{\mathcal{L}}_{0,2}^{\Sigma,1}$ *is in* $\mathrm{CDFA}(\mathtt{fin}', \subseteq) \setminus \mathrm{FA}(\mathtt{fin}, \subseteq)$.

Proposition 17. *The language* $\mathcal{L} = \Sigma_0(\tilde{\mathcal{L}}_{0,2}^{\Sigma,1} + \tilde{\mathcal{L}}_{0,3}^{\Sigma,1})$ *is in* $\mathrm{FA}(\mathtt{fin}, \subseteq)$ *but not in* $\mathrm{CDFA}(\mathtt{fin}', \subseteq)$.

Proof. We have $\mathcal{L} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}, \subseteq)}$ for the FA $\mathcal{A} = (\Sigma, \{q_0, q_1, q_2, q_3, q_4, q_5\}, T, q_0, \{\{q_2\}, \{q_4, q_5\}\})$ where T is depicted on Figure 1. For the sake of argument, assume that $\mathcal{L} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}', \subseteq)}$ for a CDFA $\mathcal{A} = (\Sigma, Q, T, q_0, \mathcal{F})$. Let $\delta : Q \to Q$ be the transition function of $\mathcal{A}$.

We first show that if u and v are two words such that u is a prefix of v starting by a 0 and $\delta(q_0, u) = \delta(q_0, v)$ then $|u|_1 = |v|_1 \pmod 6$. Let us denote $k = |u|_1 \pmod 6$ and $k' = |v|_1 \pmod 6$. If x is an ω-word, then the set of states visited finitely often by the path labeled by ux is included in the set of states visited finitely often by the path labeled by vx. Then, whenever ux is rejected for some x, vx is rejected. We take $x = 1^{(5-k)}0^\omega$ (resp. $x = 1^{(7-k)}0^\omega$), as ux is not in the language, it is rejected and vx is also rejected. We deduce that $|vx|_1 = k' + 5 - k$ (resp. $|vx|_1 = k' + 7 - k$) is congruent to 1 or 5 modulo 6. This implies that $k = k'$. Let $n = |Q|$ and $x = 010^n10^\omega$. As x is in $\mathcal{L}$, there exists $F \in \mathcal{F}$ such that $\mathtt{fin}'(p) \subseteq F$ where p is the path labeled by x. Let $S = \{q_0\} \cup \left\{\delta(q_0, x_{[0,k]}) : k \in [0, n+1]\right\}$, according to the above lemma, $S \subseteq \mathtt{fin}'(p)$. Moreover, we can find two integers $i < j$ such that $\delta(q_0, 010^i) = \delta(q_0, 010^j)$, then the path p' labeled by $y = 010^\omega$ is such that $\mathtt{run}'(p') = S$. Finally, $\mathtt{fin}'(p') \subseteq \mathtt{run}'(p') = S \subseteq \mathtt{fin}'(p) \subseteq F$ and y is recognized by $\mathcal{A}$ but $y \notin \mathcal{L}$. We get a contradiction. $\square$

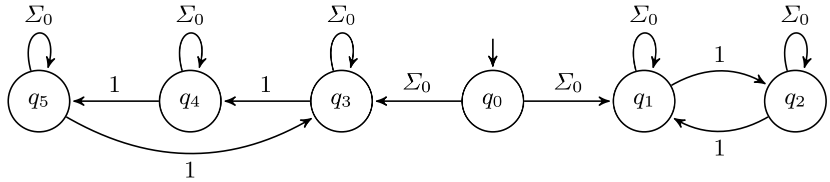

Fig. 1. A FA recognizing $\Sigma_0(\tilde{\mathcal{L}}^{\Sigma,1}_{0,2} + \tilde{\mathcal{L}}^{\Sigma,1}_{0,3})$ under the condition $(\mathtt{fin}, \subseteq)$.

Proposition 18. *The language* $\mathcal{L} = \Sigma(11\Sigma^* + 0)^\omega$ *is in* $\mathrm{FA}(\mathtt{fin}, \subseteq) \setminus \mathsf{G}_\delta$.

Proof. We have $\mathcal{L} = \mathcal{L}^{(\mathtt{fin},\subseteq)}_{\mathcal{A}}$ for the FA $\mathcal{A} = (\Sigma, \{q_0, q_1, q_2, q_3\}, T, q_0, q_1)$ where T is depicted on Figure 2. It is straightforward to prove that $\mathcal{L}$ is not in G_δ. $\square$

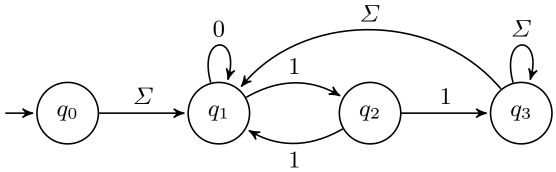

Fig. 2. A FA recognizing $\Sigma(11\Sigma^* + 0)^\omega$ under the condition $(\mathtt{fin}, \subseteq)$.

Theorem 19. *The classes induced by* $(\mathtt{fin}, \subseteq)$ *and* $(\mathtt{fin}', \subseteq)$ *satisfy the following relations:*

1. $\mathsf{F} \subsetneq \mathrm{CDFA}(\mathtt{fin}, \subseteq) \subsetneq \mathrm{CDFA}(\mathtt{fin}', \subseteq) \subsetneq \mathsf{G}_\delta$,
2. $\mathrm{CDFA}(\mathtt{fin}, \subseteq) \subsetneq \mathrm{FA}(\mathtt{fin}, \subseteq)$ *and* $\mathrm{CDFA}(\mathtt{fin}', \subseteq) \subsetneq \mathrm{FA}(\mathtt{fin}', \subseteq)$,
3. $\mathrm{FA}(\mathtt{fin}, \subseteq) \subsetneq \mathrm{FA}(\mathtt{fin}', \subseteq)$.

There is no other relation for the classes induced by $(\mathtt{fin}, \subseteq)$ *and* $(\mathtt{fin}', \subseteq)$ *except those obtained by transitivity with previously known classes.*

Proof. The inclusions of the first point comes from the Proposition 11, Corollary 5 and Proposition 12, respectively. By Propositions 13, 16 and 15, respectively, the inclusions are strict. The inclusions of the second point are clear and by Proposition 17 it is strict. The inclusions of the third point are a consequence of the Corollary 5 and by Proposition 16 they are strict.

The incomparability with the other known classes comes from Proposition 15 which proves that G and CDFA$(\mathtt{ninf}, \sqcap)$ are not subclasses of FA$(\mathtt{fin}', \subseteq)$ and from Propositions 13 and 18 which prove that CDFA$(\mathtt{fin}, \subseteq)$ is not a subclass of F_σ and FA$(\mathtt{fin}, \subseteq)$ is not a subclass of G_δ, respectively. □

7 The acceptance condition $(\mathtt{fin}, =)$ and $(\mathtt{fin}', =)$

In the previous section we have proved that the class CFA$(\mathtt{fin}, \subseteq)$ is pretty high in the hierarchy. However, it is incomparable with $\mathsf{F}^\mathsf{R}_\sigma \cap \mathsf{G}^\mathsf{R}_\delta$ and it does not contain any open language. In this section, we are going to show two more classes which have nicer properties.

Lemma 20. *Let $a, b \in \Sigma$ be two distinct letters and $\mathcal{L}$ a language such that $\mathcal{L} \cap \{a,b\}^* b^\omega = \mathcal{L}^{\{a,b\},a}_{0,2}$. If $\mathcal{L} = \mathcal{L}^{(\mathtt{fin},=)}_\mathcal{A}$ or $\mathcal{L} = \mathcal{L}^{(\mathtt{fin}',=)}_\mathcal{A}$ for a* CDFA *$\mathcal{A}$ then $\mathcal{A}$ has a loop on its initial state labeled by b^k for some $k > 0$.*

Proof. Let $\mathcal{A} = (\Sigma, Q, T, q_0, \mathcal{F})$ be a DFA such that $\mathcal{L} = \mathcal{L}^{(\mathtt{fin},=)}_\mathcal{A}$ or $\mathcal{L} = \mathcal{L}^{(\mathtt{fin}',=)}_\mathcal{A}$. For the sake of argument, assume that q_0 does not belong to a loop labeled by b's. Let δ be the transition function of $\mathcal{A}$. For all word x, denote by p_x the path in $\mathcal{A}$ labeled by the word x.

Define a sequence of integers $(k_i)_{i\in\mathbb{N}}$ such that, denoting the finite word $b^{k_0}ab^{k_1}a \dots ab^{k_i}$ by u_i, for all $i \in \mathbb{N}$, $\delta(q_0, u_i)$ does not belong to a loop labeled by b's but $\delta(q_0, u_i0)$ does. As q_0 is not on a loop labeled by b's, we define k_0 as $\max\left\{j \in \mathbb{N} : \forall j' > j, \delta(q_0, b^{j'}) \neq \delta(q_0, b^j)\right\}$. Assume that k_i is defined for some $i \in \mathbb{N}$. Then, the state $\delta(q_0, u_ia)$ does not belong to a loop labeled by b's. Indeed, otherwise the words $x = u_ib^\omega$ and $y = u_iab^\omega$ verify $\mathtt{fin}_\mathcal{A}(p_x) = \mathtt{fin}_\mathcal{A}(p_y)$ and $\mathtt{fin}'_\mathcal{A}(p_x) = \mathtt{fin}'_\mathcal{A}(p_y)$ (in both cases, the states which appear in those sets are states reached by reading u_i in $\mathcal{A}$ counting or not the first state). This is not possible because only one of this words is accepted by $\mathcal{A}$. We define k_{i+1} as $\max\left\{j \in \mathbb{N} : \forall j' > j, \delta(q_0, u_i10^{j'}) \neq \delta(q_0, u_i10^j)\right\}$.

Since Q is finite, there exists $i < j$ such that $\delta(q_0, u_i) = \delta(q_0, u_j)$. The words $x = u_jb^\omega$ and $y = u_jab^\omega$ verify $\mathtt{fin}_\mathcal{A}(p_x) = \mathtt{fin}_\mathcal{A}(p_y)$ and $\mathtt{fin}'_\mathcal{A}(p_x) = \mathtt{fin}'_\mathcal{A}(p_y)$ (see Figure 3) but as above only one of these words is accepted by $\mathcal{A}$. We get a contradiction. □

Proposition 21. *The language $\mathcal{L}^{\Sigma,1}_{0,2}$ is in* CDFA$(\mathtt{fin}', =) \smallsetminus$ CDFA$(\mathtt{fin}, =)$.

Proof. We have $\mathcal{L}^{\Sigma,1}_{0,2} = \mathcal{L}^{(\mathtt{fin}',=)}_\mathcal{A}$ for the CDFA $\mathcal{A} = (\Sigma, \{q_0, q_1, q_2\}, T, q_0, \{\emptyset, \{q_0, q_1\}\})$ where $(p, a, q) \in T$ if and only if $a \neq 1$ and $p = q$ or $a = 1$

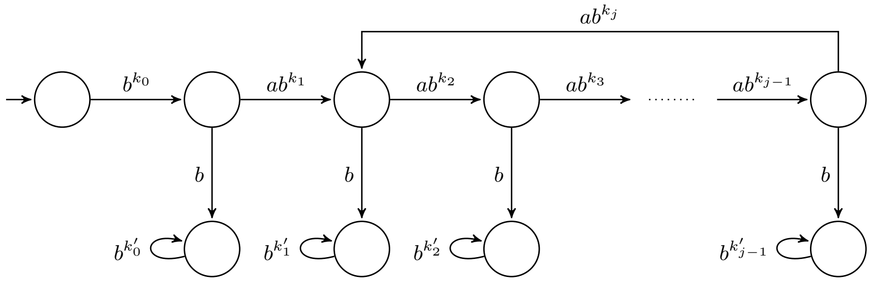

Fig. 3. A figure illustrating the construction in Lemma 20 with $i = 1$.

and $(p, q) \in \{(q_0, q_1), (q_1, q_2), (q_2, q_1)\}$. If $\mathcal{L}_{0,2}^{\Sigma,1}$ would be recognized by a CDFA $\mathcal{B}$ under condition $(\mathtt{fin}, =)$, $\mathcal{B}$ could be assumed normalized by Lemma 4. But as $\mathcal{L}_{0,2}^{\Sigma,1} \cap \{0,1\}^* \, 0^\omega = \mathcal{L}_{0,2}^{\{0,1\},1}$, by Lemma 20, this automaton should have a loop on its initial state. This is not possible and $\mathcal{L}_{0,2}^{\Sigma,1}$ is not in CDFA$(\mathtt{fin}, =)$. □

Proposition 22. *The language* $\mathcal{L} = \mathcal{L}_{0,2}^{\Sigma,0} + \mathcal{L}_{0,2}^{\Sigma,1}$ *is not in* CDFA$(\mathtt{fin}', =)$.

Proof. For the sake of argument, assume that $\mathcal{L} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}', =)}$ for a CDFA $\mathcal{A} = (\Sigma, Q, T, q_0, F)$. As $\mathcal{L} \cap \{0,1\}^* \, 0^\omega = \mathcal{L}_{0,2}^{\Sigma,1}$, by Lemma 20, there exists k such that there exists a loop on q_0 labeled by 0^k. Symmetrically, there exists k' such that there exists a loop on q_0 labeled by $1^{k'}$. As $0^\omega \in \mathcal{L}$, $\emptyset \in \mathcal{F}$. The path p labeled by $x = (0^k 1^{k'})^\omega$ verifies $\mathtt{fin}'(p) = \emptyset \in \mathcal{F}$. Then x is recognized but x is not in $\mathcal{L}$. We have a contradiction. □

Remark 23. Using similar methods as in the proof of Lemma 20 and Proposition 22, we can prove that the language $\Sigma(\mathcal{L}_{1,2}^{\Sigma,0} + \mathcal{L}_{1,2}^{\Sigma,1})$ is not in CDFA$(\mathtt{fin}', =)$. Since CDFA$(\mathtt{fin}', =)$ is clearly closed under complementation, $\Sigma(\tilde{\mathcal{L}}_{0,2}^{\Sigma,0} \cap \tilde{\mathcal{L}}_{0,2}^{\Sigma,1})$ is not in CDFA$(\mathtt{fin}', =)$.

Proposition 24. *The language* $\mathcal{L} = \mathcal{L}_{1,2}^{\Sigma,0} + \mathcal{L}_{1,2}^{\Sigma,1}$ *is in* CFA$(\mathtt{fin}, \sqcap)$ *but not in* CDFA$(\mathtt{fin}', =)$.

Proof. By Proposition 9 and using the non-determinism, it is clear that $\mathcal{L}$ is in CFA$(\mathtt{fin}, \sqcap)$. By Remark 23, $\mathcal{L} \notin$ CDFA$(\mathtt{fin}', =)$. □

Proposition 25. *The language* $\mathcal{L} = \Sigma(\tilde{\mathcal{L}}_{0,2}^{\Sigma,0} \cap \tilde{\mathcal{L}}_{0,2}^{\Sigma,1})$ *is in* FA$(\mathtt{fin}, \subseteq)$ *but not in* CDFA$(\mathtt{fin}', =)$.

Proof. We have $\mathcal{L} = \mathcal{L}_{\mathcal{A}}^{(\mathtt{fin}, \subseteq)}$ for the CFA $\mathcal{A} = (\Sigma, \{q_0, q_1, q_2, q_3, q_4, q_5, q_6, q_7\}, T, q_0, \{q_2, q_3, q_4, q_6\})$ where T is depicted in Figure 4 (here $\bar{\Sigma}$ means $\Sigma \setminus \{0, 1\}$). This automaton is split in two disjoint parts. A path which visits the state q_5 is successful if and only if q_5 (and then q_7) is visited an infinite number of times, if and only if its label contains an infinite number of occurrences of the pattern 01, if and only if its label contains in infinite number of a's and b's.

A path visiting q_1 is successful if and only if q_1 is visited an infinite number of times. Let p be a successful path visiting q_1, let ax be its label where $a \in \Sigma$ and

$x \in \Sigma^\omega$. If $|x|_0$ (resp. $|x|_1$) is finite, the set $\mathtt{inf}_\mathcal{A}(p)$ is included in $\{q_1, q_2\}$ or in $\{q_3, q_4\}$ (resp., in $\{q_1, q_3\}$ or in $\{q_2, q_4\}$). Since p is successful, q_1 is in $\mathtt{inf}_\mathcal{A}(p)$, therefore $\mathtt{inf}_\mathcal{A}(p)$ is included in $\{q_1, q_2\}$ (resp., in $\{q_1, q_3\}$) and $|x|_0$ (resp., $|x|_1$) is even. The converse is clear. By Remark 23, $\mathcal{L} \notin \mathrm{CDFA}(\mathtt{fin}', =)$. □

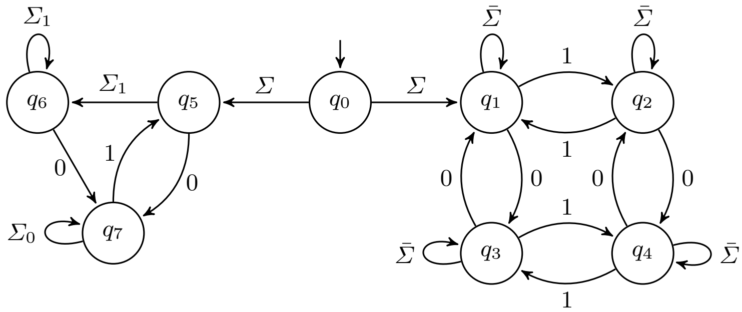

Fig. 4. A FA recognizing $\Sigma(\tilde{\mathcal{L}}_{0,2}^{\Sigma,0} \cap \tilde{\mathcal{L}}_{0,2}^{\Sigma,1})$ under the condition $(\mathtt{fin}, \subseteq)$.

8 Conclusions

This paper is a step further in the study of the hierarchy of ω-languages induced by accepting conditions found in the literature. Figure 5 illustrates the hierarchy and highlights the contribution of this paper.

This research can be continued along several directions. First of all, some inclusions of classes induced by $(\mathtt{ninf}, \sqcap)$ into $\mathrm{CDFA}(\mathtt{fin}', =)$ are still open.

Secondly, in [2], the authors proved that a slight generalization of classical Büchi result: all second order definable accepting conditions induce ω-rational languages. It would be very interesting to study what is the impact of weaker fragments of logic over the classification provided here.

Another promising research direction considers the closure properties of the newly found classes of ω-languages.

Finally, the decidability of the new classes is certainly a promising research direction.

References

1. Julius Richard Büchi. Symposium on decision problems: On a decision method in restricted second order arithmetic. In Patrick Suppes Ernest Nagel and Alfred Tarski, editors, *Logic, Methodology and Philosophy of Science Proceeding of the 1960 International Congress*, volume 44 of *Studies in Logic and the Foundations of Mathematics*, pages 1–11. Elsevier, 1960.
2. Julien Cervelle, Alberto Dennunzio, Enrico Formenti, and Julien Provillard. Acceptance conditions for ω-languages and the Borel hierarchy. Preprint, arXiv:1310.5032 [cs.FM], 2013. Submitted.

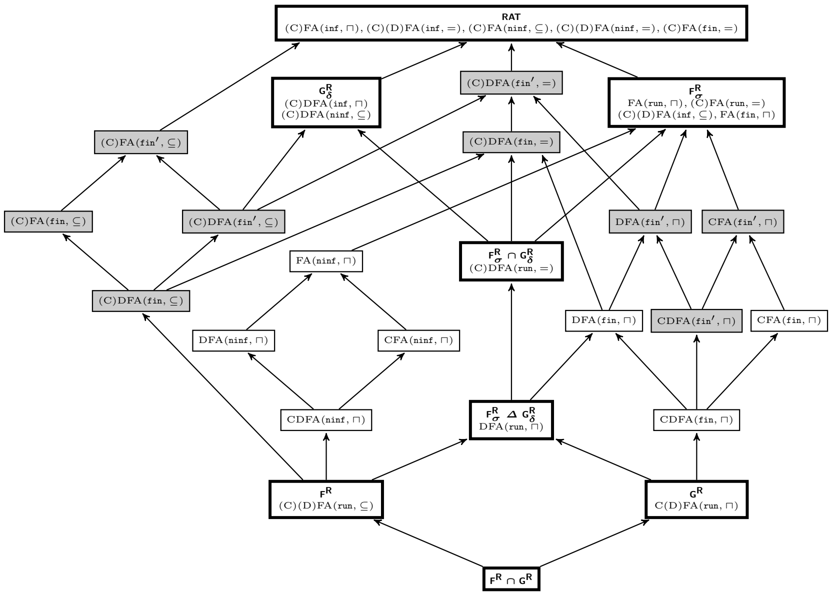

Fig. 5. The hierarchy of classes of ω-languages. The Borel hierarchy appears in bold boxes. Grayed boxes show the new classes studied in this paper. Arrows represent inclusions between classes. Classes in the same box are equal. The possibly missing arrows are from classes induced by $(\mathtt{ninf}, \sqcap)$ to CDFA$(\mathtt{fin}', =)$, the question is open in this case.

3. Alberto Dennunzio, Enrico Formenti, and Julien Provillard. Acceptance conditions for ω-languages. In *Developments in Language Theory*, pages 320–331, 2012.
4. Juris Hartmanis and Richard E. Stearns. Sets of numbers defined by finite automata. *American Mathematical Monthly*, 74:539–542, 1967.
5. Laurence H. Landweber. Decision problems for omega-automata. *Mathematical Systems Theory*, 3(4):376–384, 1969.
6. Igor Litovsky and Ludwig Staiger. Finite acceptance of infinite words. *Theoretical Computer Science*, 174(1-2):1–21, 1997.
7. Tetsuo Moriya and Hideki Yamasaki. Accepting conditions for automata on ω-languages. *Theoretical Computer Science*, 61:137–147, 1988.
8. David E. Muller. Infinite sequences and finite machines. In *Proceedings of the 1963 Proceedings of the Fourth Annual Symposium on Switching Circuit Theory and Logical Design*, SWCT '63, pages 3–16. IEEE Computer Society, 1963.
9. Dominique Perrin and Jean-Eric Pin. *Infinite words, automata, semigroups, logic and games*, volume 141 of *Pure and Applied Mathematics*. Elsevier, 2004.
10. Ludwig Staiger. ω-languages. In *Handbook of formal languages*, volume 3, pages 339–387. 1997.
11. Ludwig Staiger and Klaus W. Wagner. Automatentheoretische und automatenfreie charakterisierungen topologischer klassen regulärer folgenmengen. *Elektronische Informationsverarbeitung und Kybernetik*, 10(7):379–392, 1974.
12. Klaus W. Wagner. On ω-regular sets. *Information and Control*, 43(2):123–177, 1979.